

AD-A132 313

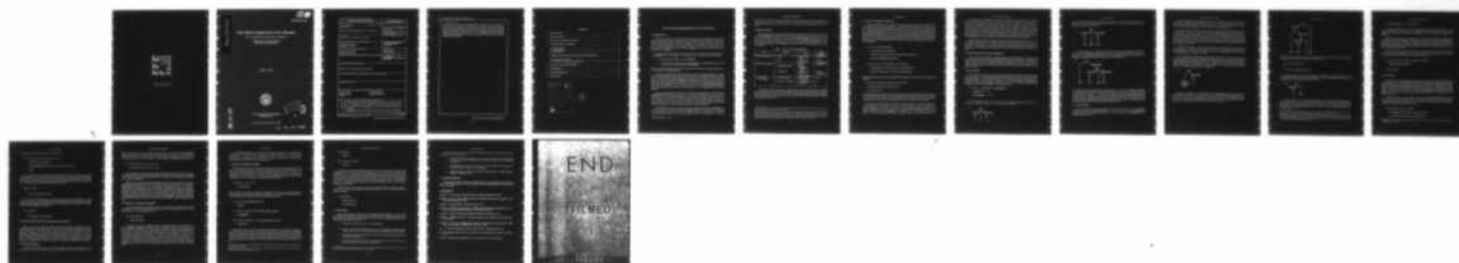
NOUN PHRASE COMPRESSION IN NAVY MESSAGES(U) NAVAL
RESEARCH LAB WASHINGTON DC J C BACHENKO ET AL.
31 AUG 83 NRL-8748

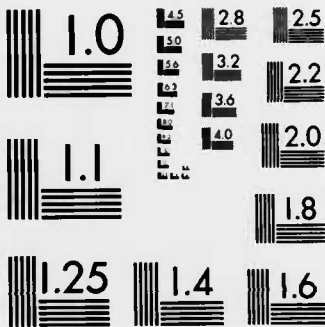
1/1

UNCLASSIFIED

F/G 5/7

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

50

NRL Report 8748

ADA132313

Noun Phrase Compression in Navy Messages

JOAN C. BACHENKO AND CONSTANCE L. HEITMEYER

*Computer Science and Systems Branch
Information Technology Division*

August 31, 1983



NAVAL RESEARCH LABORATORY
Washington, D.C.

DTIC
ELECTE
SEP 09 1983
S D E

DTIC FILE COPY

Approved for public release; distribution unlimited.

83 09 07 028

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NRL Report 8748	2. GOVT ACCESSION NO. AD-A132318	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) NOUN PHRASE COMPRESSION IN NAVY MESSAGES		5. TYPE OF REPORT & PERIOD COVERED Interim report on a continuing NRL problem
		6. PERFORMING ORG. REPORT NUMBER 7590-134:JCB:CLH
7. AUTHOR(s) Joan C. Bachenko and Constance L. Heitmeyer		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Research Laboratory Washington, DC 20375		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61153N RR014-09-42 75-0207-0-3
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, DC 20360		12. REPORT DATE August 31, 1983
		13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Military messages Message automation Grammar Artificial intelligence Computational linguistics		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) An important goal of the Navy's future message systems is to be able to interpret automatically the English narrative in Navy messages. To accomplish this, such systems must incorporate a natural language parser that determines the structure of English phrases and sentences. The parser does this by using a grammar , a set of rules that describes all acceptable patterns in the language. A central aim of our research is to develop a grammar that is oriented toward the particular style, or sublanguage , used in Navy messages. In this report, we describe a partial grammar (Continued)		

20. ABSTRACT (Continued)

for telegraphic constructs in the message sublanguage, i.e. constructs in which words that would normally occur are omitted when the meaning of the phrase is clear without them. Specifically, we are concerned with telegraphic noun phrases of the form *Modifier + Noun + Modifier*, where either or both of the modifiers are optional. We present a grammatical description of **complex acronyms** like CINCLANT ('Commander-in-Chief for the Atlantic'), focusing on the rules that assign these expressions an internal structure. Given this description, we show that complex acronyms are hybrid expressions in that they share the properties of single words like CNO and of phrases like *arrival Norfolk*.

CONTENTS

1. INTRODUCTION	1
2. THREE NP TYPES	2
3. LEXICAL VS COMPLEX ACRONYMS	3
4. THE STRUCTURE OF COMPLEX	4
4.1 Syntactics Rules	4
4.2 Lexical Properties	5
4.3 Noun Classes	8
5. COMPLEX ACRONYMS AND TELEGRAPHIC NOUN STRINGS	9
5.1 Word-order Restrictions	9
5.2 Restrictions on Complement Interpretation	10
6. COMPLEX ACRONYMS AS WORDS	11
7. CONCLUSIONS	12
8. ACKNOWLEDGMENTS	13
9. REFERENCES	13

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	



NOUN PHRASE COMPRESSION IN NAVY MESSAGES

1. INTRODUCTION

To increase the effectiveness of military decision-makers and to save manpower, the Navy is gradually replacing many manual procedures for handling messages with computer-based message systems (Heitmeyer and Wilson, 1980). An important future goal of such systems is that they be able to interpret automatically the English narrative in Navy messages. One component of these systems, therefore, is a natural language parser that determines the structure of English phrases and sentences. The parser determines what these structures are by using a **grammar**, a set of rules that describes all acceptable patterns in the language.

A central aim of our research is to develop a grammar that is oriented toward the particular style, or **sublanguage**, used in Navy messages. To accomplish this goal, we need to answer two questions:

- What are the basic properties of the sublanguage?
- To what extent do these properties reflect properties of standard English and to what extent are these properties special to the sublanguage?

Once they are identified, these properties in conjunction with a grammar of standard English can be used to develop the **sublanguage grammar**.

One important aspect of the message sublanguage is its use of "telegraphic constructs," i.e. phrases in which words that would normally occur are omitted when they are not needed for semantic interpretation. The basic properties of such phrases are determined by the kind of words they omit. Telegraphic constructs exhibit **syntactic compression** by omitting elements whose role can be determined solely by the grammatical context; an example is the noun phrase *remainder month* which omits the semantically empty preposition *of* and the definite article *the*. Constructs exhibit **discourse compression** when the omitted element is a content word whose existence can be grammatically determined but whose identity must be inferred from the discourse context, e.g., the missing subject in *will arrive tomorrow*.

This report examines syntactic compression in a large class of sublanguage noun phrases, where a **noun phrase**, or **NP**, consists of a head noun and one or more modifiers. For example, *the sponsor's arrival in Norfolk* is a NP in which the head noun is *arrival* and the modifiers are *the sponsor's* and *in Norfolk*. *Sponsor arrival Norfolk* is a syntactically compressed NP in which the modifiers are simply *sponsor* and *Norfolk*. As we will show, syntactic compression in sublanguage NPs reflects a syntax that is simpler and a semantics that is more restricted than that of ordinary English because certain linguistic options are excluded, e.g., the rules that generate definite articles and prepositions.

The purpose of this report is to present a subset of the rules that constitute the sublanguage grammar. We describe three types of NP, focusing on **complex acronyms** like CINCLANT ('Commander in Chief for the Atlantic') and CINCLANTFLT ('Commander in Chief for the Atlantic Fleet'). Examples of each type are given in section 2 of this report, where we outline a simple taxonomy for the NP data

that concern us. Sections 3 and 4 present linguistic rules for the complex acronyms, including the syntactic rules that determine their internal structure. Some differences between complex acronyms and other NPs are discussed in sections 5 and 6. Section 7 presents our conclusions and plans for future work.

2. THREE NP TYPES

We are concerned with three types of sublanguage NPs. Examples of each type are listed in Table 1. The **lexical acronyms** in (1a-b) are single nouns; though synonymous with structured phrases, they have no internal structure themselves. In contrast, each **complex acronym** in (2a-b) consists of more than one noun and has an internal structure similar to that of the **telegraphic noun strings** in (3a-b). Both lexical acronyms and complex acronyms are names that usually refer to an organization or to both a job title and an organization.¹ The telegraphic noun strings in (3a-b) are compressed NPs with no special reference.²

Table 1 — Three Types of Sublanguage NPs.

Type	Example	Synonym	Semantic Category
(1) Lexical Acronyms	a. ASO	Aviation Supply Office	organization
	b. CNO	Chief of Naval Operations	job title/ organization
(2) Complex Acronyms	a. HUMRESMANCEN	Human Resource Management Center	organization
	b. CINCLANT	Commander in Chief for the Atlantic	job title/ organization
(3) Telegraphic Noun Strings	a. remainder month	the remainder of the month	n/a
	b. sponsor arrival Norfolk	the sponsor's arrival in Norfolk	n/a

The following shows that complex acronyms follow the basic structural patterns of telegraphic noun strings but with fewer variations: the relationships among the nouns that comprise a complex acronym are limited and fixed by a small number of rules, making these NPs internally rigid and, hence, similar to words. This and their use as names account for the resemblance between the complex acronyms and lexical acronyms.

¹Complex acronyms that end in INST or NOTE refer to document titles, e.g. CINCLANTFLTINST ('Commander- in-Chief of the Atlantic Fleet Instruction') and CINCLANTFLTNOTE.

²All well-formed examples cited in this report come from a collection of approximately 1600 Navy messages and from the lists of sublanguage expressions assembled in Bachenko and Parker (1981), The Plain Language Address Directory (1981), and Wedertz (1979). As these sources show, acronyms occur very frequently in the sublanguage. A typical example of how they are used in message narrative is REQUEST NAVSURFLANT READSUPPGRU NORVA DET SKED TECH ASSIST (= *We request Naval Surface Forces for the Atlantic, Readiness Support Group, Norfolk, Virginia, Detachment schedule technical assistance*).

3. LEXICAL VS COMPLEX ACRONYMS

Linguists have generally viewed acronyms as an aberrant word type having little, if anything, to do with serious questions of linguistic description. Aronoff (1976:20) gives us a recent example of this view, discussing acronyms like NATO and blends like *smog* (*smoke* + *fog*) under the heading 'Oddities'. Aronoff observes that acronyms like NATO have no internal structure because their formation is not governed by grammatical rules. They are irregular forms in the language and therefore odd. But Aronoff's description applies only to what we have called the lexical acronyms; it fails to hold for complex acronyms like CINCLANT, which have an internal structure whose properties are determined by syntactic rules. We shall thus adopt Aronoff's proposals in part and assume that lexical acronyms like those in example (4) and word abbreviations such as those in example (5) are single nouns. Synonyms are in parentheses.

- (4) a. ASO ('Aviation Supply Office')
 b. CNO ('Chief of Naval Operations')
 c. DONPIC ('Department of the Navy Program Information Center')
 d. SOPA ('Senior Officer Present Afloat')
- (5) a. NAV ('Navy'), b. CEN ('center'), c. GRU ('group'),
 d. RON ('squadron'), e. HUM ('human'), f. LANT ('Atlantic')
 g. SURF ('surface forces') h. ASW ('antisubmarine warfare')

Because their description is simple and straightforward, we shall have little more to say about the lexical acronyms.

The complex acronyms can be treated in one of two ways:

- as single nouns, like the lexical acronyms, or
- as NP constructions.

In the first approach, each complex acronym would be listed in the lexicon along with its meaning. The second approach assumes that the short words, word abbreviations, and lexical acronyms that constitute a complex acronym can be listed in the lexicon together with their properties; syntactic rules determine a structure for the complex acronym and semantic rules use the structure to derive a meaning.

We shall take the second approach and claim that complex acronyms are not single nouns but syntactic constructions whose structure is governed by a subset of the rules that form such NPs as *remainder month* and *arrival Norfolk*. They have an order of constituents in which an obligatory noun, called the **head**, may be preceded by an optional NP modifier called the **specifier** and followed by an optional NP modifier called the **complement**. In NAVSURFLANT ('Naval Surface Forces for the Atlantic'), for example, the head noun is SURF, the specifier is NAV, and the complement is LANT.

As we will show, our approach to the description of complex acronyms reveals aspects of syntactic compression in NPs that are obscured by the 'single noun' approach. Our approach also has the advantage of economy and flexibility. Since most complex acronyms are combinations of a small number of nouns, it is more efficient to list only these nouns in the lexicon rather than complete acronyms; in either case, the rules that combine the nouns are needed for the description of NPs in general. (These rules are presented in section 4 of this report.) Moreover, the meaning of a complex acronym can involve several items of information, depending on the nouns that make up the acronym and the way they are combined. Our approach suggests a systematic way of deriving complete meanings for acronyms based on their structure; including the complete meaning in a single lexical entry would be awkward and difficult since basic generalizations about acronym meaning could not be expressed.

Finally, complex acronyms are a dynamic class: new ones are created and used all the time. For the 'single word' approach, this is a problem since each new acronym requires a new lexical entry. In our approach, introduction of new complex acronyms is less troublesome, since new acronyms are described by existing linguistic rules and since most new acronyms are composed of nouns that are already present in the lexicon; a new lexical entry is needed only when the acronym includes a new noun.

4. THE STRUCTURE OF COMPLEX ACRONYMS

We assume that the formation of syntactic constructions, including NPs, is governed by two interacting subsystems of the grammar: **syntactic rules** that determine the structure of a phrase and a **lexicon** that lists individual words and their properties (Chomsky 1957, 1970, 1981). In this section, we present the general rules of NP formation that determine a structure for complex acronyms. We then discuss the lexical properties of the individual words, e.g., CINC and LANT, that are combined to form complex acronyms.

4.1 Syntactic Rules

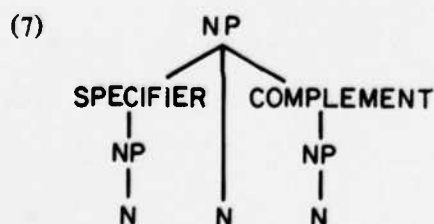
The syntactic rules in (6a-c), an adaptation of the NP formation rules of Chomsky (1970), assign an internal structure to complex acronyms and other NPs. These are phrase structure rules whose purpose in the grammar is to identify abstract categories (e.g., NP) and to specify a left-right order and hierarchical relations for the categories (Chomsky 1957). The parentheses mean that Specifier and Complement are optional. The category N, where N stands for noun, is obligatory because syntactic structures must have a head and the head of a NP is always a noun.

(6) a. NP ----> (SPECIFIER) N (COMPLEMENT)

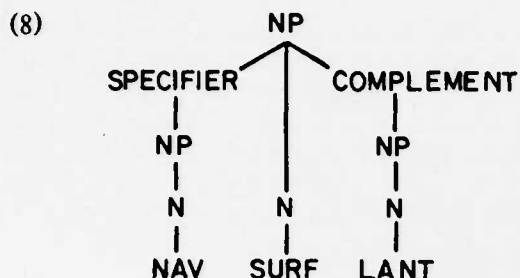
b. SPECIFIER ----> NP

c. COMPLEMENT ----> NP

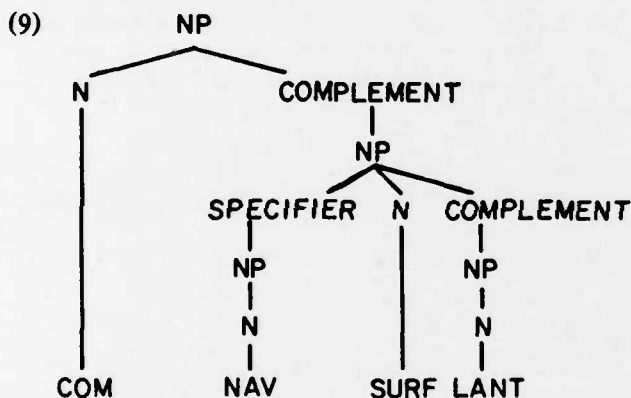
The tree diagram in example (7) is an example of a NP structure assigned by rules (6a-c); the Specifier and Complement options of (6a) are both present in example (7).



A syntactic description is complete when words are assigned to those parts of the structure that are marked by N. For instance, one realization of the structure in (7) is the syntactic description in (8), which is assigned to NAVSURFLANT.



The rules given in (6a-c) are recursive. Hence a NP structure may stand alone as a complex acronym or it may be embedded in a larger complex acronym. For example, the NP structure in (7) can stand alone as the acronym NAVSURFLANT or it may be embedded, as shown in example (9), as a complement in COMNAVSURFLANT ('Commander of NAVSURFLANT').



In such cases, the head of the construction is the 'highest N', i.e., the N that is immediately dominated by the highest NP. Thus COM is the head of the construction in example (9) and SURF is the head of the construction in example (8). It is the head of the construction that enforces agreement restrictions such as that between a subject and a predicate. For example, the sentence *CINCLANT believes that this program is useful* is acceptable because the predicate *believes that this program is useful* agrees with the head noun *CINC*. The sentence *CINCLANT was formed millions of years ago* is unacceptable because the predicate agrees with the complement *LANT* instead of the head.

4.2 Lexical Properties

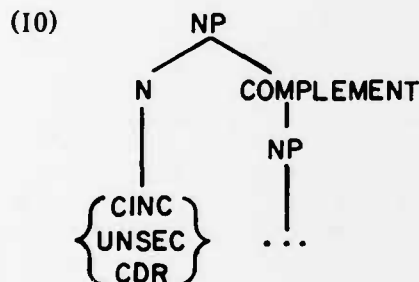
Nouns and other content words are listed in the lexicon along with a set of word properties. These include the word's **category membership**, e.g., noun, verb, adjective; the **subcategorization features** that describe the contextual restrictions on a word's occurrence; and **selection features** that describe semantic constraints.

Category membership and subcategorization features are examples of syntactic properties, i.e., properties that describe some aspect of syntactic structure. Thus nouns are marked with the category feature [N], which signifies membership in the category 'Noun'. Subcategorization features use symbols such as 'Specifier' and 'Complement' to describe structural contexts for head nouns.

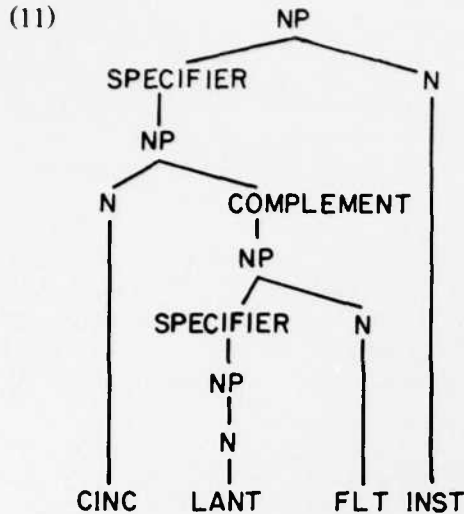
Selection features are examples of semantic properties; they describe the semantic restrictions on word occurrence. For instance, selection features govern the agreement between a subject and a predicate and between a head and its modifiers. In the example *CINCLANT believes that this program is useful*, CINCLANT agrees with the predicate because *believe* is a verb that selects human nouns as subjects and because the head noun CINC is a human noun. Similarly, in the phrase *sincere person* the head noun *person* agrees with its modifier *sincere* because *sincere* requires a human noun and *person* satisfies the requirement; the expression *sincere rock* is unacceptable because *rock* is nonhuman and therefore fails to agree with its modifier.

Although we will continue to refer to selection features, our main concern is with the syntax of complex acronyms and, hence, the syntactic properties of nouns. Our discussion in this section focuses on subcategorization. We describe three syntactic classes of nouns, illustrated by CINC, INST, and LANT, that subcategorization allows us to identify.

When CINC, UNSEC ('Undersecretary'), and CDR ('Commander') occur in a NP structure, they are never preceded by a specifier. These nouns always appear as the first word of an acronym and they are always followed by a complement: UNSECNV and CINCLANT are both possible acronyms but UNSECCINCLANT and CINCUNSECNV are not. More generally, they only occur as head of a right-branching structure like example (10). (For convenience, we have omitted some details from (10); the '...' refers to any expansion of the embedded NP.)



Such facts about CINC, etc., are summarized by the subcategorization feature [0_Complement], which states that no specifier can precede the noun and that a NP complement must follow it; that is, the noun can only occur in the position of head in example (10), where no word precedes the head. A structure such as (10) may occur independently, as with CINCLANTFLT, or it may be embedded as a specifier, e.g., CINCLANTFLTINST, which has the structure in example (11):

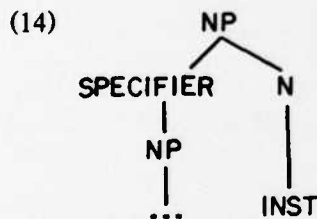


Each of the nouns in example (12) is therefore described in the lexicon by the collection of category, subcategorization, and selection features in the lexical entry (13).³

(12) CINC, UNSEC, CDR

(13) [N], [0__Complement], [job title]

In contrast, the noun INST is always preceded by a specifier and is always the final noun; thus SECNAVINST ('Secretary of the Navy Instruction') is an acronym but INSTSECNAV is not. This noun only occurs as the head noun in structures like (14).



The lexicon accounts for such facts by assigning INST the subcategorization feature [Specifier_0], which restricts the noun to the position of head in example (14). The presence of Specifier in the feature means that INST must be preceded by a NP; the presence of 0 means that no word can occur to the right of the noun. INST thus appears in the lexicon with the category, subcategorization, and selection features given in (15):⁴

³The selection features described in this example and the other examples in this section are incomplete. A more complete account of example (13), for example, would exclude certain noun sequences, e.g., CINCCINCLANT.

⁴Although correct in general, a more restricted description of INST may be necessary. The subcategorization feature in example (15) only requires that INST occur as an acronym-final head; example (15) does not prevent structures with INST as head from being embedded as a complement in a larger structure. It may thus be necessary to add additional information to example (15) if, as we suspect, INST cannot be embedded in a larger structure.

(15) INST [N], [Specifier__0], [document title]

Place names like LANT and PAC form an entirely separate class. Unlike CINC, INST, etc., these nouns, which are called 'locatives', can never be the head of an acronym construction. They can only appear in the position of the complement, as in SUBLANT, or in the position of the specifier, as in LANTFLT.

In the model of grammar we are assuming, only nouns that can occur as heads of a phrase are marked for subcategorization features; i.e., the grammar assumes that only these nouns are subject to restrictions on their context. Modifiers are not subject to contextual constraints and therefore have no subcategorization features; their role instead is to impose constraints on the head noun in a way that is similar to those discussed with our earlier example *sincere person*, where a selection constraint on context is imposed by *sincere* on *person*.

The locative nouns in example (16), therefore, can be entered into the lexicon as (17). Because (17) contains no subcategorization features, these nouns will not occur as heads. The feature [locative] is a semantic property that allows the word to be used as a place name.

(16) LANT, PAC, EUR ('Europe'), SIXTHFLT ('Sixth Fleet'),
EIGHT ('Eighth Naval District')

(17) [N], [locative]

4.3 Noun Classes

As previously noted, subcategorization refers solely to syntactic context and noun classes based on subcategorization are syntactic classes; for example, CINC, UNSEC, and CDR all belong to a single syntactic class described by the subcategorization feature [0__Complement]. Selection features such as [job title] and [locative] describe semantic classes, which may or may not correspond to syntactic classes. For example, CINC, UNSEC, CDR, and REP ('Representative') all share the same semantic property, [job title], but do not belong to the same syntactic class because REP, unlike CINC and the others, can never occur in acronym-initial position; LANTFLTREP is a possible acronym, but REPLANTFLT is not.

In some cases, however, selection features do pair up with particular subcategorization features so that semantic and syntactic classes correspond. This happens with SEC and COM, which are homographs whose meanings vary according to the syntactic context. When they are acronym-initial, as in examples (18a-c), SEC and COM mean 'job title':

- (18) a. SECDEF ('Secretary of Defense')
b. COMSUBGRU ('Commander of Submarine Group')
c. COMNAVDIST ('Commandant of Naval District')⁵

⁵Although 'job title' is the general meaning of acronym-initial COM, the noun is ambiguous between two specific meanings: 'Commander' and 'Commandant'.

In any other context, they have the meaning 'organization', e.g.,

- (19) a. NAVSEC ('Navy Ships Engineering Center')
- b. PACOM ('Pacific Command')
- c. LANTCOMOPSUPPFAC ('Atlantic Command Operations Support Facility')

SEC and COM follow these patterns because each belongs to two semantic classes that correspond to two syntactic classes. More precisely, each is listed in the lexicon with the collection of information in (20) and (21). By pairing the feature [job title] with the subcategorization feature [0__Complement], (20) allows SEC and COM to mean 'job title' only when they occur as the first noun of an acronym. This puts them in the same class as CINC, UNSEC, and CDR.

- (20) a. SEC, COM
- b. [N], [0__Complement], [job title]

In (21), the feature [organization] is paired with the subcategorization feature [Specifier__], which allows the noun to occur in any position but acronym-initial. COM or SEC preceded by a specifier therefore will always have the meaning 'organization'. This puts them into a class that is both semantically and syntactically different from the first.

- (21) a. SEC, COM
- b. [N], [Specifier__], [organization]

5. COMPLEX ACRONYMS AND TELEGRAPHIC NOUN STRINGS

Section 4 of this report outlines a syntactic system in which acronyms are assigned an internal structure by the general rules of NP formation in (6a-c) and by a lexicon that includes such words as CINC, COM, GR11, etc. Given this system, we claim that complex acronyms constitute a type of NP analogous to telegraphic noun strings like *remainder month*: both are NP structures of the form (*Specifier*) *Noun* (*Complement*) and both exhibit syntactic compression. Our concern here and in the section that follows is with those properties of complex acronyms that distinguish them as a separate group. In this section, we describe two ways in which the complex acronyms differ from telegraphic noun strings: they permit fewer variations in word order and complement interpretation is restricted.

5.1 Word-order Restrictions

Complex acronyms exhibit little variation in form because acronym syntax includes few of the options that allow for structural complexity, and hence variability, in NPs. Word-order patterns, for

example, are fixed by rules (6a-c). They cannot be changed by the syntactic rules, called **transformations**, that allow nouns to be moved within a syntactic structure. Thus acronyms never exhibit word order variations like those in examples (22a-b). In these telegraphic noun strings, a NP transformation allows *sponsor* to occur to the left or the right of the head without a corresponding change in meaning.

(22) a. sponsor arrival ('the sponsor's arrival')

b. arrival sponsor ('the arrival of the sponsor')

These telegraphic noun strings share the same basic meaning because they share the same underlying structure, *NP+N*. This structure directly describes *sponsor arrival*. In *arrival sponsor*, it is changed to *N + NP* by the syntactic transformation 'Move NP', which permits NPs that follow the head noun to be interpreted as NP Specifiers.

Because syntactic transformations are excluded from acronym formation, there can be no NP displacement within an acronym construction. Thus complex acronyms of the form *NP + N* will have no predictable variants of the form *N + NP*, and vice versa. For example, since the *NP + N* structures CRUDESGRU ('Cruiser/Destroyer Group') and NRLINST ('Instruction for NRL') are acronyms, it does not follow that GRUCRUDES ('Group of Cruisers and Destroyers') and INSTNRL ('Instruction for NRL') are also acronyms (in fact, the latter two forms do not exist). This also explains why NAVSEC and SECNAV are distinct forms. If word-order patterns are fixed by rules (6a-c), then the *N + Complement* construction SECNAV cannot be a variant of the *Specifier + N* construction NAVSEC and the lexical properties of SEC that we discussed earlier will determine a different meaning for each form.

5.2 Restrictions on Complement Interpretation

Lexical simplifications affect variability in other ways. Unlike other telegraphic NPs, complex acronyms make no semantic distinctions among NP complements. Thus contrasts like those in examples (23a-b) do not occur in complex acronyms:

(23) a. departure Norfolk

b. destruction Norfolk

In example (23a), *Norfolk* is a **locative** complement that describes *where*. In example (23b), *Norfolk* is an **objective** complement that describes *what*. Thus *Norfolk* in example (23a) is interpreted as *from Norfolk*, but in example (23b) it is interpreted as *of Norfolk*. To describe these distinctions, the lexical entry for nouns like *departure* include a selection feature that requires the complement to be a locative noun. This feature acts as a filter on the interpretation. If the complement is a locative noun, as in *departure Norfolk*, then the complement type is locative. If the complement is not a locative noun, as in *departure sponsor*, then it must be treated as a displaced specifier and interpreted as objective. In contrast, the lexical entries for nouns like *destruction* do not include a locative selection feature and consequently the complements of these nouns receive the default interpretation 'objective' when the complement is a locative noun. The rules that describe these selection features are given in Fitzpatrick (1983).

A complement that appears in a complex acronym always is an objective noun; for example, the locative noun PAC ('Pacific') is always interpreted as an objective complement, as in CINCPAC and NAVSURFPAC. Thus nouns that take complements in complex acronyms, e.g., CINC and SURF, always require the complement to be an objective noun. Like the noun *destruction*, the lexical entries for these nouns do not include the locative selection feature.

6. COMPLEX ACRONYMS AS WORDS

We noted earlier in this report that complex acronyms are hybrid constructions sharing the properties of both words and phrases. On the one hand, they possess the internal structure of NPs; on the other, they resemble words because, unlike the telegraphic noun strings, they do not preserve the boundaries between individual nouns in a structure. These boundaries are erased by a contraction rule that applies to a string of nouns, e.g., example (24a), and collapses the string into a single unit like example (24b). The '+' sign in example (24a) designates word boundaries.

(24) a. CINC + LANT + FLT

b. CINCLANTFLT

When the letters on both sides of a boundary are identical, the contraction rule assimilates them into a single occurrence. The acronyms in (b) of examples (25) to (27), for instance, are derived from the corresponding string in (a) by the contraction rule and letter assimilation.⁶

(25) a. PAC + COM ('Pacific Command')

b. PACOM

(26) a. COM + MINE + RON ('Commander of Mine Squadron')

b. COMINERON

(27) a. AIR + AS + SUPP + U ('Air Antisubmarine Support Unit')

b. AIRASUPPU

Because they have no internal word boundaries, complex acronyms, after contraction, appear and behave as words. For example, they act as single words with respect to the attachment of inflectional elements.⁷ In English, these elements can only occur outside a word boundary--they are never inserted into a word. The complex acronym in example (28a) therefore has the plural form in example (28b). Though similar in structure to *Secretary of the Navy*, SECNAV lacks the internal word boundaries that allow plural formation by attachment of *s* to the head; example (29a) is an acceptable form but (29b) is not:

⁶There is at least one exception, namely, COMMATWING ('Commander of Medium Attack Wing'), where the lack of assimilation appears to be idiosyncratic.

⁷Elements like the *s* of plural nouns and the *ed* of past tense verbs are inflectional elements; the noun *arrivals* and the verb *departed* are inflections of *arrival* and *depart*, respectively.

(28) a. SECNAV

b. SECNAV_s

(29) a. Secretaries of the Navy

b. SECsNAV

The absence of internal word boundaries can also account for the exclusion of semantically empty prepositions (e.g., *of* in *gallons of fuel*) from complex acronyms. Such prepositions are attached in noun strings of the form $N + NP$; they are a type of inflectional element since the rule that attaches them must have access to both phrase and word boundaries. They lack semantic content because their role, as well as their form, is determined by the lexical properties of a head noun and a NP complement (Levi 1979). In telegraphic noun strings, therefore, these prepositions are optional. For example, *of* may or may not occur in the noun string *arrival sponsor*; the preposition *to* may or may not occur in *opposition current allotments*.⁸

When the NP is a complex acronym, however, the preposition cannot be inserted because acronyms have no internal word boundaries. The preposition therefore is not an option. None of the following is acceptable:

(30) a. SECofNAV

b. CINCoFLANTFLT

c. NAVSURFforLANT

7. CONCLUSIONS

A parser that interprets the narrative in Navy messages must use a grammar, i.e., a set of rules that describes acceptable patterns in the narrative. In this report, we have described a subset of these rules, namely, the syntactic rules that determine the form of complex acronyms and telegraphic noun strings. Specifically, we have shown that:

- acronyms are divided into two types: lexical and complex;
- complex acronyms like CINCLANTFLT form a class of telegraphic NP whose syntactic structure parallels that of such noun strings as *gallons fuel* and *opposition raised allotments*,
- both complex acronyms and telegraphic noun strings can be described by general rules of NP formation in English; and
- only minor modifications to these rules are needed to account for the properties that distinguish complex acronyms from other telegraphic NPs.

⁸Not all prepositions are semantically empty. In the noun string *arrival by helicopter*, the preposition cannot be omitted without changing the complement type of *helicopter* from instrumental to objective.

Future work will be aimed at developing a more general description of telegraphic phrases in the sublanguage. This work will focus on:

- extending the description of selection (i.e., semantic) restrictions that hold between nouns in a NP structure and formulating rules that use these restrictions, e.g., in complement interpretation;
- comparing syntactic compression in NPs, e.g., *sponsor arrival Norfolk*, with compression in other phrase types, e.g., *sponsor arrived Norfolk*; and
- formulating syntactic rules that will recognize *Norfolk destroyed* as a passive sentence, analogous to *Norfolk's destruction*.

8. ACKNOWLEDGMENTS

We are grateful to Eileen Fitzpatrick and Judith Froscher for their thoughtful review of this report and for many helpful suggestions. We also thank Karen Jensen and Virginia Teller for their comments on an earlier version.

9. REFERENCES

- Aronoff, M., *Word Formation in Generative Grammar*. MIT Press Cambridge, MA 1976.
- Bachenko, J. and R. A. Parker, *Reference Manual: Navy Message Words and Word Frequencies*. NRL Memorandum Report 4673, 1981.
- Chomsky, N., *Syntactic Structures*. The Hague: Mouton, 1957.
- Chomsky, N., "Remarks on Nominalization" in *Readings in English Transformational Grammar*, ed. by R. Jacobs and Peter Rosenbaum, Waltham, MA, 1970 Ginn, pp. 184-221.
- Chomsky, N., *Lectures on Government and Binding*. Dordrecht, Foris Publications, 1981.
- Fitzpatrick, E., "Binding Conventions in Generative Grammar," Ph.D. Dissertation, New York University, NY, 1983.
- Heitmeyer, C. and S. Wilson, "Military Message Systems: Current Status and Future Directions. *IEEE Trans. on Communications*. COM-28, No. 9, 1645-1654, (1980).
- Levi, J., *The Syntax and Semantics of Complex Nominals*. New York: Academic Press, 1979.
- 1981 Plain Language Address Directory, Commander, Naval Telecommunications Command. Washington, D. C.
- Wedertz, B., *Dictionary of Naval Abbreviations*, 1979 Annapolis, MD: Naval Institute Press.

END

FILMED

9-83

DTIC